Eur J Forest Res (2012) 131:1761–1770
DOI 10.1007/s10342-012-0630-9

ORIGINAL PAPER

DNA sequence variation and development of SNP markers in beech (*Fagus sylvatica* L.)

S. Seifert · B. Vornam · R. Finkeldey

Received: 23 September 2011 / Revised: 10 February 2012 / Accepted: 16 March 2012 / Published online: 3 April 2012

Abstract European beech (*Fagus sylvatica* L.) is one of the most important deciduous tree species in Central Europe. The potential of beech to adapt to climate change, higher temperatures, and less precipitation in the summer months is still unknown. Most studies in beech used microsatellite, AFLP (amplified fragment length polymorphism), or isozyme markers, which have only a restricted potential to analyze adaptation. Only few studies investigated genes probably involved in the adaptation to drought stress and bud phenology in beech. In this study, SNP (single nucleotide polymorphisms) markers were developed in order to analyze adaptation and their technical advantages compared to microsatellites and AFLPs were discussed. Partial sequences of ten candidate genes probably involved in drought stress and/or bud phenology were identified at the genomic level, and SNPs and indels (insertions/deletions) in coding and non-coding regions were analyzed. Plant material was sampled along a precipitation gradient in Germany. In total, 8,145 bp were sequenced and analyzed, 4,038 bp were located in exon and 4,107 bp in intron regions. 63 SNPs and 11 indels were detected, which are differently distributed over the studied gene regions. The nucleotide diversity ranged from 0 to 6.62 ($\pi \times 10^{-3}$) and is comparable to other tree species, whereas the mean nucleotide diversity (2.64) for *F. sylvatica* is comparatively low. These results will help to investigate the genetic basis of drought stress and bud burst and to conduct association mapping in natural populations. Furthermore, the detected SNPs can also be used for population genetic studies.

Keywords Climate change · Adaptation · Candidate gene · *Fagus sylvatica* · SNPs

Communicated by R. Matyssek.

Electronic supplementary material The online version of this article (doi:10.1007/s10342-012-0630-9) contains supplementary material, which is available to authorized users.

S. Seifert (✉) · B. Vornam · R. Finkeldey
Faculty of Forest Sciences and Forest Ecology,
Büsgen Institute, Forest Genetics and Forest Tree Breeding,
Georg-August-University Göttingen, Büsgenweg 2,
37077 Göttingen, Germany
e-mail: sseifer@gwdg.de

Introduction

Forest tree populations consist of sessile, long-lived organisms which must survive temporally varying environmental conditions that are presently also affected by accelerated global climate change. Hence, the presence and maintenance of genetic variation at genes controlling adaptive traits is important for the long-term persistence and stability of forest tree populations in order to survive heterogeneous conditions.

Genetic markers are ultimately based on the variation of DNA sequences. However, the sequences of the currently most commonly used genetic markers in beech (*Fagus sylvatica* L.) are not directly observed and are usually unknown. Only particular aspects of the variation are investigated within PCR-amplified DNA fragments such as the number of tandem repeats in microsatellite motives (e.g., Pastorelli et al. 2003) or the absence or presence of restriction sites in amplified fragment length polymorphisms (AFLPs; e.g., Gailing and von Wühlisch 2004). The amplified genomic regions are usually either unknown such as in anonymous AFLPs or are located in non-coding regions of the DNA (most microsatellites). Accordingly, most of the variation at molecular DNA-based markers is

assumed to be selectively neutral. In addition, the accurate scoring of microsatellite markers (SSRs) and AFLPs can be difficult due to PCR and electrophoresis artifacts, and the comparability between different laboratories is problematic. Although different SSR loci can be multiplexed for a higher throughput, the multiplex process is complex, not always successful, and limited to a low number of loci.

Furthermore, isozymes are also important markers. These biochemical markers can be used to assess genetic diversity at gene loci coding for enzymes, which serve important functions in the metabolism of plants. For beech, the genetic diversity has been studied extensively at selected isozyme gene loci (e.g., Müller-Starck and Ziehe 1991; Müller-Starck and Starke 1993). However, the analysis of isozymes allows to explore only a fraction of the underlying sequence variation, and only few gene loci coding for selected soluble enzymes can be investigated by means of enzyme electrophoresis. Furthermore, it is questionable whether isozymes are suitable to detect adaptive variation or if most of the markers are neutral (e.g., Eriksson 1998 and references therein).

Comparative sequencing is the ultimate method to detect variation within any DNA fragment. Today, it is already possible to analyze and to compare whole genomes of organisms using high-throughput sequencing technologies, also called next-generation sequencing. The most frequently used techniques at the moment are 454 (Roche), SOLiD (Life Technologies), and Illumina (e.g., reviewed by Glenn 2011, Deschamps and Campbell 2010). However, this method is still too expensive to analyze a sufficient number of individuals for population genetic studies or the study of adaptation in natural populations. For non-model organisms like trees, where most of the genomes are not sequenced yet, next-generation sequencing is not an established technique. Considering these limitations, the most promising markers for the study of adaptation at the moment are SNPs (single nucleotide poylmorphism), that is, the substitutions of only one nucleobase. SNPs are the most frequent variations found in DNA (Brookes 1999), and the analysis is not restricted to special enzymes. SNP marker can, unlike isozymes, also be used to analyze regions controlling the transcription of genes, for example, transcription factor binding sites. In comparison to SSRs and AFLPs, they are valuable markers to study adaptation of plants, for example, to changing environmental conditions (Gailing et al. 2009). For human and plant model organisms, this type of marker is already establishment and often used (e.g., *Populus tremula*, Ingvarsson 2004). However, SNP markers are nowadays more and more applied for non-model organisms like most of the forest trees (e.g., Seeb et al. 2011, Helyar et al. 2011).

Unfortunately, SNP analyses in human populations revealed that only few SNPs can be associated with phenotypic traits (Yoshiura et al. 2006). Some of these SNPs with a direct impact on phenotypes are likely to be under selection, while the vast majority of SNPs are likely to behave selectively neutral. However, besides the study of adaptation, SNPs in non-coding regions can also be used instead or additional to other neutral markers. The analysis of an unprecedented number of mostly selectively neutral SNP loci allows new insights in the population genetic structure of species that cannot be found with other genetic markers. For example, the observation of more than 500,000 SNPs in over 3,000 Europeans revealed overall genetic differentiation patterns among humans on the continent closely resembling their spatial distribution on the continent (Novembre et al. 2008). Furthermore, comparing AFLP and microsatellite markers with SNP markers, the latter markers have some important advantages. The scoring is unambiguously and comparable between laboratories, even if different platforms are used for the analysis. Jones et al. (2007) compared SSR and SNP markers in maize and concluded that SNP markers have a lower level of missing data and are more reliable. For the analysis of SNPs, multiplexing can be conducted easily, and thus, the throughput is very high. Estimations show that SNP costs are lower in comparison to SSR markers (Jones et al. 2007). However, efficiency and costs strongly depend on the platform used for SNP scoring, but it is predictable that the efficiency costs will decrease in the future.

The aim of this study is to detect SNPs within candidate genes, related to phenotypic traits in beech. European beech (*Fagus sylvatica* L.) is one of the predominant and most important tree species in European forests and covers a large geographic range in Central Europe. The species is wind-pollinated, predominantly outcrossing, a monoecious tree with heavy fruits and therefore with limited seed dispersal.

So far, in beech, most studies on the genetic diversity and differentiation were focused on the spatial genetic structure or on the impact of different silvicultural treatments using AFLP and microsatellite markers (Vornam et al. 2004; Buiteveld et al. 2007; Nyári 2010; Oddou-Muratorio et al. 2011). In the context of global climatic changes, predicting less precipitation in summer and higher precipitation in winter contradictory opinions exist whether beech will be adaptable to the enhanced drought stress conditions in the summer months (Gessler et al. 2007; Rennenberg et al. 2004; Ammer et al. 2005). Another effect of the predicted global change is the extending growing season influencing the growth of beech in the future. Earlier bud burst is supposable, which will lead to an increasing risk of late frost damage. The analysis of the variation within 'candidate' genes potentially involved in adaptation to a phenotypic trait is one possibility to investigate the genetic background of adaptation. Until now, only few studies aim to identify genes that are

involved in drought stress response and bud phenology in beech (Lalagüe et al. 2010), and only a limited number of beech sequences are available (Jimenez et al. 2008; Olbrich et al. 2005, 2010; Schlink 2011). Therefore, the candidate gene approach described here is based on both published *F. sylvatica* sequences and orthologous sequences identified in other plant species such as oaks (Gailing et al. 2009; Vornam et al. 2011).

SNPs were analyzed in both coding (exons) and non-coding regions (introns) of the identified genes. For the purpose of using SNP markers additionally or in place of microsatellite markers, it is necessary to analyze both regions. For the study of adaptation, SNPs in coding regions changing the amino acid composition of the gene products (non-synonymous SNPs) are most interesting, but non-coding regions can also be of relevance. Whereas non-synonymous SNPs potentially lead to changes of protein structures, SNPs in intron regions potentially influence gene splicing and enable a single gene to increase its coding capacity producing several structurally distinct isoforms (Baek et al. 2008).

The results described here are a prerequisite for association mapping in natural populations in order to identify SNPs correlated to phenotypic traits like drought stress response and bud phenology. Other applications of the analysis of SNPs are, for example, population genetic studies concerning the history, structure and demography of populations or molecular systematic studies and parentage analyses (Garvin et al. 2010; Morin et al. 2004). The SNPs identified in this study are suitable for population genetic investigations complementing other frequently used markers such as microsatellites and AFLPs. Furthermore, this study provides the first estimates of nucleotide and haplotype diversity in *F. sylvatica*.

Materials and methods

Plant material

Fresh leaves were sampled in early summer 2009 in three different regions of northern Germany along a rainfall gradient (Table 1). All stands are jointly investigated by several research groups within the collaborative project 'Climate Impact and Adaptation Research in Lower Saxony' (KLIFF; http://www.kliff-niedersachsen.de.vweb5-test.gwdg.de/?page_id=26). Each region is represented by two populations differing in their soil type. Three trees per population were used for SNP identification. Thus, the total sample size was 3 (regions) × 2 (populations/region) × 3 (trees/population) = 18 trees. The investigated trees were separated by a distance of at least 50 m to minimize the risk of sampling related plants (Vornam et al. 2004).

Selection of candidate genes

All candidate genes have been chosen based on literature surveys suggesting an impact of the genes on either drought stress or bud phenology (Table 2). The Evoltree EST database (http://www.evoltree.org) and the EMBL Nucleotide Sequence Database (http://www.ebi.ac.uk/embl/) were mainly used to find corresponding *F. sylvatica* sequences. Alternatively, sequences of *Quercus petraea* were transferred to *F. sylvatica* (Vornam et al. 2007; Vidalis 2011). The selected sequences were verified by a TBLASTX search (Washington University Basic Local Alignment Search Tool Version 2.0) and used for primer design in order to amplify the corresponding genomic regions in beech.

DNA isolation, amplification, cloning, and sequencing

Total DNA was extracted from leaves using the DNeasy™ 96 Plant Kit (Qiagen, Hilden, Germany). The amount and the quality of the DNA were analyzed by 0.8 % agarose gel electrophoresis with 1 × TAE as running buffer (Sambrook et al. 1989). DNA was stained with ethidium bromide, visualized by UV illumination, and compared to a Lambda DNA size marker (Roche).

Primers for amplification and direct sequencing of the amplification product (Table 3) were designed by using the program Primer3 (v.0.4.0; Rozen and Skaletsky 2000; http://frodo.wi.mit.edu/). Primers were checked for self-annealing, dimer, and hairpin formations using the

Table 1 Sampling sites in Germany, Lower Saxony, and Saxony-Anhalt

Closest village	Position	Elevation (m)	Annual amount of precipitation[a] (mm)	Amount of precipitation from April to September[a] (mm)	Soil type
Calvörde	N52 22.819 E11 17.406	65	543	294	Sand
	N52 24.238 E11 15.661	75	543	294	Loam
Göhrde	N53 08.660 E10 52.003	90	665	347	Sand
	N53 07.379 E10 49.224	85	675	349	Loam
Unterlüß	N52 49.831 E10 18.985	130	766	374	Sand
	N52 49.894 E10 19.183	130	766	374	Loam

[a] Provided by National Climate Monitoring of Deutscher Wetterdienst (DWD)

Table 2 Selected candidate genes related to drought stress response or bud phenology

Name (abbreviation)	Gene	Drought stress/bud phenology	Reference with investigated species
ALDH	*Aldehyde dehydrogenase*	Drought stress	Gao and Han (2009) (*Oryza sativa*), Guo et al. (2009) (*Hordeum vulgare*), Sathyan et al. (2005) (*Pinus halepensis*)
Cry	*Cryptochrome*	Bud phenology	Muleo et al. (2001) (*Prunus cerasifera*)
Dhn	*Dehydrin*	Drought stress and bud phenology	Beck et al. (2007) (*Cicer pinnatifidum*), Gonzáles-Martínez et al. (2006) (*P. taeda*), Jimenez et al. (2008) (*F. sylvatica*), Ramanjalu and Bartels (2002) (*Picea glauca*), Wachowiak et al. (2009) (*Pinus sylvestris*), Vornam et al. (2011) (*Q. petraea*)
ERD	*Early response to dehydration*	Drought stress	Eveno et al. (2008) (*Pinus pinaster*), Street et al. (2006) (*Populus* spp.)
IDH	*Isocitrate dehydrogenase*	Drought stress	Liu et al. (2010) (*Zea mays*)
APX1, APX4, GPX	*Peroxidases (ascorbate and glutathione)*	Drought stress	Lu et al. (2010) (*Zea mays*), Street et al. 2006 (*Populus* spp.)
PhyB	*Phytochrome B*	Drought stress and bud phenology	Boggs et al. (2010) (*Arabidopsis thaliana*), Frewen et al. (2000) (*Populus* spp.), Ingvarsson et al. 2006 (*P. tremula*)
CHZFP	*Cys-his-zinc finger protein*	Drought stress	Lu et al. (2010) (*Zea mays*), Street et al. (2006) (*Populus* spp.)

program Oligo calc: Oligonucleotide Properties Calculator (http://www.basic.northwestern.edu/biotools/oligocalc.html). PCR amplifications were conducted in a 15 µl volume containing 2 µl of genomic DNA (about 10 ng), 7.5 µl HotStar*Taq* Master Mix Kit (Qiagen, Hilden, Germany), and 0.3 µM of each forward and reverse primer. The PCR protocol consisted of an initial denaturation step of 95 °C for 15 min, followed by 35 cycles of 94 °C for 60 s (denaturation), different temperatures according to the primers (Table 3) for 45 s (annealing), 72 °C for 90 s (extension), and a final extension step of 72 °C for 20 min.

PCR products were analyzed by 1 % agarose gel electrophoresis with 1 × TAE as running buffer (Sambrook et al. 1989). DNA was stained with ethidium bromide and visualized by UV illumination. PCR products were excised from gel and purified using the Geneclean® kit (MP Biomedicals, Illkirch, France). The purified products were cloned into a pCR2.1 vector using the TOPO TA Cloning® kit (Invitrogen, Carlsbad, CA) with slight modifications. The inserts were amplified by colony PCR using M13 forward (-20) (5′-GTAAAACGACGGCCAG-3′) and M13 reverse (5′-CAGGAAACAGCTATGAC-3′) primers, visualized by agarose gel electrophoresis, excised from the gel and purified (see above). Three to four different clones of the fragments were sequenced using both M13 forward and M13 reverse primers in order to identify the presence of different haplotypes within individuals (heterozygotes) and to control for sequencing errors. The sequencing reaction was carried out with the Big Dye® Terminator v.3.1. Cycle Sequencing Kit (Applied Biosystems) based on the dideoxy-mediated chain termination method (Sanger et al. 1977). Sequencing reactions were run on an ABI 3100*xl* Genetic Analyser (Applied Biosystems). The sequenced fragments were verified by a TBLASTX search. Putative introns and exons were determined following the GT-AG rules (Breathnach et al. 1978).

Data analysis

For editing and visual examination of the sequences as well as for the analysis of SNPs and indels (insertions/deletions) within the genes, the sequences were aligned using Codon Code Aligner (CodonCode cooperation, http://www.codoncode.com) and BioEdit version 7.0.9.0 (Hall 1999) using ClustalW multiple alignment (Thompson et al. 1994). Only polymorphisms with Phred scores above 25 in the chromatograms were considered (Ewing et al. 1998). Only SNPs appearing at least twice were analyzed in order to avoid sequencing errors. Haplotype diversity, nucleotide diversity (π), and F_{ST} values were calculated excluding indels using DnaSP v.5.0 (Librado and Rozas 2009).

Results

Fragments from ten different genes were successfully amplified, identified, and analyzed. After sequencing, all fragments were verified using TBLASTX search. Any similarity with an *E* Value of less than 10^{-3} was considered to be a hit. In total, 9,468 bp were analyzed with 4,418 bp in exon regions and 5,050 bp in intron regions (Table 4). All exons and introns could be determined following the GT-AG rule. No alternative splicing was found. The reading frame was assessed according to the TBLASTX results (see above).

Table 3 Primer sequences and corresponding annealing temperatures for the selected candidate genes (Accession No: EMBL Nucleotide Sequence Database (http://www.ebi.ac.uk/embl/))

Name	Gene	Primer sequence (5′-3′)	Annealing temperature (°C)	Accession no
ALDH	*Aldehyde dehydrogenase*	F: AAG ATC TGG TGT TGA AAA TGG AG	53	FR774766
		R: TGC ATT CTT CAA AGG AGT GAC		
APX1[b]	*Ascorbate peroxidase*	F Part 1: AGG CGA AGA GAA AGC TCA GG	55	FR774767
		R Part 1: AAG AAA GCA ACT ATC AGC CTC A		
		F Part 2: AAG CAG ATT TGT TGA CAT TAA TAT TTC		
		R Part 2: GCA AAG AAG GCA TCC TCA TC		
APX4 (Part 1)[a]	*Ascorbate peroxidase*	F Part 1: ATC AAG GGA ACG CTT TCT ACG	55	FR775801
		R Part 1: TCC ACA TCA CAT CTC AAC AGC		
APX4 (Part 2)[a]	*Ascorbate peroxidase*	F Part 2: GGC CTC TTA AGT GCC AAT TC	55	FR775801
		R Part 2: CTC CCC TCT GGA TCT GGT TC		
Cry[b]	*Cryptochrome*	F Part 1: CTT GAG ATG ATG CTC TTG GTT G	53	FR775802
		R Part 1: ATG GGC TCA ATT GGA GAA TC		
		F Part 2: TTT TCT CCA CAG GGA TCA CG		
		R Part2: AAG TCA TGC TTG GGA CCA TC		
Dhn	*Dehydrin*	F: TGC ACC CCA AAA TGA TGA AT	54	FR772355
		R: TGA TCC CCT TCT TCT CAT GG		
ERD	*Early response to dehydration*	F: GGC AAT GGA CGT AAT TTC TCA	51	FR775803
		R: CTG GGC TGC TGA ATC GTC		
GPX	*Glutathione peroxidase*	F: GGC TGC CAT GCC TTT CTC	55	FR796394
		R: GAA ATC ATA GAT AGT CTT CTC CGT AGC		
IDH (Part 1)[a]	*Isocitrate dehydrogenase*	F: GTG ATC AGT ACA GGG CAA CTG	50	FR796392
		R: AAG GTA CAA GAG GGG CTT TG		
IDH (Part 2)[a]	*Isocitrate dehydrogenase*	F: GAT GAT ATG GTT GCT TAT GCC ATG	50	FR796392
		R: GGT TTC ACC ACC TTT CTG ATG GAC		
PhyB	*Phytochrome B*	F: CAG GCA TCG AGG TTT TTG TT	50	FR774765
		R: GAA GGG AAT GCA CCT AGC AG		
CHZFP	*Cys-his-zinc finger protein*	F: CTT TGC AAG GAT GAG ACT GG	50	FR796395
		R: ACG CAT CTG ATG AGC ATT TG		

[a] Both parts belong to the same gene but the sequenced parts do not overlap

[b] Both parts belong to the same gene and the two parts overlap

Insertions/deletions

In seven different genes, 11 indels (insertions/deletions) were identified, mainly in intron regions (Table 4). Some of them showed a microsatellite repeat motif (see supplementary material). Only two indels also represented by microsatellite motives were found within coding regions (gene *ERD* and *CHZFP*). The lengths of these indels were multiples of 3 bps; thus, the reading frame is not shifted.

Single nucleotide polymorphisms

Single nucleotide polymorphisms only appearing once were excluded from the analyses in order to avoid the selection of false positives caused by sequencing errors, although they could be true SNPs. Therefore, only common SNPs are presented here that may be also present in *F. sylvatica* trees in other regions in Europe than investigated in this study. Considering these limitations, in total, 63 SNPs were found differently distributed over the analyzed gene fragments. The results indicate that numerous of these SNPs are linked (see supplementary material). Excluding the potentially linked SNPs from the analysis, 45 SNPs remain. However, because of the low number of investigated trees, the linkage of these SNPs is not unambiguous and it is not possible to clearly define a set of tag SNPs.

More SNPs were found in non-coding regions (1 SNP every 112 bp) than in coding regions (1 SNP every 245 bp). Eighteen SNPs were found in coding regions, and seven of them were non-synonymous. All non-synonymous SNPs led to an amino acid exchange, no one caused an early stop codon. The number of haplotypes ranged from

Table 4 Length, exons, introns, indels, and SNPs of the amplified candidate genes

Gene name	Total length (bp)	No and length (bp) of exons	No and lengths (bp) of introns	No indels	No SNPs	No of non-coding SNPs	No of synonymous SNPs	No of non-synonymous SNPs
ALDH	519	3/350	2/169	1	8	4	2	2
APX1	1,566	7/577	7/989	4	15	12	3	
APX4 (Part 1)	608	2/107	3/501	1	6	6	0	0
APX4 (Part 2)	814	2/238	3/576	1	11	9	1	1
Cry	1,439	2/379	1/1,060	1	1	1	0	0
Dhn	546	2/455	1/91	0	3	1	0	2
ERD (Part 1)	317	1/242	1/75	0	2	2	0	0
ERD (Part 2)	152	1/145	1/7	1	0	0	0	0
GPX	224	1/224	0/0	0	1	0	0	1
IDH (Part 1)	646	3/259	3/387	0	10	8	2	0
IDH (Part 2)	474	3/222	2/252	1	4	2	1	1
PhyB	301	1/301	0/0	0	2	0	2	0
CHZFP	539	1/539	0/0	1	0	0	0	0
Total	8,145	29/4,038	25/4,107	11	63	45	11	7

one to eleven. The nucleotide diversity (π) was higher at non-coding sites than at coding sites for most genes. Exceptions are the genes *GPX* and *PhyB* for which the investigated non-coding regions were very short (Table 5). Furthermore, the nucleotide diversity at synonymous sites was in most cases higher than at non-synonymous sites (Table 5).

F_{ST} was analyzed grouping the studied trees according to their region (Calvörde, Göhrde or Unterlüß), each region includes trees from two different populations. The detected values were rather low, between 0 and 0.157 with a mean value of 0.012 (Table 5). This mean value is comparable to the results of a study analyzing the same populations with nine microsatellite markers (F_{ST}: 0.022; Seifert, unpublished). However, the strongest differentiation with microsatellite loci was 0.032, whereas some candidate genes showed a considerable higher differentiation. The highest differentiations were found investigating the genes *ALDH*, *ERD* (Part1), and *IDH* (Part 2) with values above 0.05 (Table 5). Derory et al. (2010) found comparable results for SNPs analyzed in candidate genes and microsatellites for *Q. petraea*.

The partial sequence encoding aldehyde dehydrogenase was found to be of special interest. All but one of the detected SNPs (non-coding, synonymous, and non-synonymous) were represented in three different haplotypes. The indel found in the non-coding region is also linked to two other non-coding SNPs. Within this gene fragment, two non-synonymous SNPs were identified in the same codon, which were not linked. Therefore, three different amino acids are encoded depending on the combination of the SNPs indicating different lineages of the alleles. The *dehydrin* sequence is also interesting because the larger part of the sequence represents an exon region in which two SNPs were detected. Both of them are non-synonymous, and one is linked to the third non-coding SNP.

The position of the SNPs in the gene regions and additional information about the composition of the indels can be found in the supplementary material.

Discussion

In this study, parts of ten different candidate genes have been investigated. Because of the limited sequence information for *F. sylvatica*, it was not possible to sequence whole genes. However, this study was able to detect numerous SNPs and indels in non-coding and, probably more important, in coding regions of genes potentially involved in drought stress response and bud phenology. Most of the indels were found in intron regions. Only two were located in exon regions. Indels in exon regions are important due to their influence on protein structures and thus, on phenotypic trait changes (for example, reviewed by Li et al. 2002). However, short indels, like the ones that were found in this study, seem to have only minor impact on protein structures (Kim and Guo 2010). Because SNPs appearing only once were excluded from the analysis, the presented data most likely underestimate the number of SNPs. Other reasons for underestimating the number of SNPs are the limited number of investigated samples and sequencing of only three to four clones, which does not allow to correctly identify all heterozygotes. Nevertheless, 63 reliable SNPs were found. As expected, more SNPs were found in non-coding regions than in coding regions and the nucleotide diversity was higher in non-coding sites than in coding sites.

Table 5 Haplotype diversity and nucleotide diversity for the different genes (syn.: synonymous)

Gene name	No of haplotypes	Haplotype diversity	Total Nucleotide diversity ($\pi \times 10^{-3}$)	Nucleotide diversity ($\pi \times 10^{-3}$)				F_{ST}
				Non-coding sites	Coding sites	Syn. sites	Non-syn. sites	
ALDH	4	0.613	5.91	10.59	3.82	8.60	2.36	0.157
*APX*1	11	0.584	2.61	3.42	1.30	5.42	0.00	0.017
*APX*4 (Part 1)	7	0.756	3.91	4.81	0.00	0.00	0.00	−0.039
*APX*4 (Part 2)	5	0.756	3.94	4.58	2.41	4.82	1.61	0.016
Cry	2	0.157	0.11	0.15	0.00	0.00	0.00	−0.091
Dhn	3	0.627	1.83	2.70	1.66	0.00	2.16	0.044
ERD (Part 1)	3	0.629	2.52	10.67	0.00	0.00	0.00	0.058
ERD (Part 2)	1	0.000	0.00	0.00	0.00	0.00	0.00	0.000
GPX	2	0.157	0.70	0.00	0.71	0.00	0.95	0.005
IDH (Part 1)	11	0.883	6.62	8.41	4.00	17.7	0.00	−0.006
IDH (Part 2)	5	0.744	3.56	3.70	3.40	9.65	1.45	0.062
PhyB	3	0.656	2.65	–	2.65	11.83	0.00	−0.027
CHZFP	1	0.000	0.00	0.00	0.00	0.00	0.00	−0.043

Some of the non-synonymous SNPs detected in this study are of special interest because they might have an influence on the protein structure and protein function. For example, one non-synonymous SNP found in the partial sequence encoding aldehyde dehydrogenase is coding for proline, which leads to confirmation changes of the protein (Chou and Fasman 1974). The first non-synonymous SNP found in the partial *dehydrin* gene sequence leads to an amino acid substitution from aspartic acid to histidine implicating also a changed charge profile of the different genotypes from negatively charged to positively charged.

The nucleotide diversity ($\pi \times 10^{-3}$) found in this study ranged from 0 to 6.62 and is comparable to the nucleotide diversities analyzed in other tree species, for example, *Q. petraea* (1.09–14.7, Derory et al. 2010; 3.02–11.96, Gailing et al. 2009), *Quercus crispula* (6.67–7.21, Quang et al. 2008), *P. tremula* (2.7–18.8, Ingvarsson 2004), *Pinus taeda* (0.1–11.79, González-Martínez et al. 2006), and *Pseudotsuga menziesii* (2.37–13.78, Krutovsky and Neale 2005). The mean nucleotide diversity of 2.64 for *F. sylvatica* is comparatively low (*Q. petraea*: 6.15 or 5.42; *Q. crispula*: 6.93; *P. tremula*: 11.1; *P. taeda*: 7.5; *P. menziesii*: 6.55). One reason for lower nucleotide diversity values may be the exclusion of all SNPs appearing only once from the analysis (see above). However, the significance of mean values for nucleotide diversity depends on the analyzed candidate gene. Olson et al. (2010) also found in *Populus balsamifera* that the diversity is affected by the functional classification of the analyzed candidate genes. They found higher diversity in gene fragments with insertion/deletion length variation (indels) than in fragments that did not contain indels. Studies that do not include regions with length variation may slightly underestimate the overall level of nucleotide diversity.

The variation found in this study can be used to develop SNP markers and to apply them additionally or instead of neutral SSR or AFLP marker. SNP markers are more optimal markers for many applications because they are suitable for high-throughput analysis, inexpensive, highly reproducible, easy to score, comparable between different laboratories, and some SNPs clearly show higher differentiation values. Although SNP markers have some advantages, there are also some drawbacks that have to be discussed. A disadvantage of SNPs is their normally biallelic character. Thus, they are less polymorphic than SSRs. To replace ten to twenty highly polymorphic SSR markers, around 100 neutral SNP markers are necessary (Kalinowski 2002). However, the virtually unlimited number of SNP markers in the different parts of the genome of higher organisms creates opportunities for the investigation of genetic variation within species with numerous applications in population genomics. Furthermore, an ascertainment bias can occur, that is, the deviation from the expected allele frequency distribution for the case that the SNPs are identified based on only a few individuals and later used for the genotyping of a large sample set. This problem can be overcome, for example, by a direct correction of the statistical estimators (Helyar et al. 2011).

Another important application of SNPs is the study of adaptation. The SNPs found in this study can be useful, for example, to extend the investigation of Kraj and Sztorc (2009) who analyzed the variability of phenological forms (bud burst) in beech using a set of five microsatellite markers. They pointed out that the neutral microsatellite loci are not directly linked with adaptive genetic variation and the genetic differences between the phenological forms of beech (early-, intermediate-, and late-flushing individuals) have therefore no direct effect on the fitness of these forms. But genetic diversity and fitness are the basis for the ability of

forest tree populations to adapt to changes of the environment (Krutovsky and Neale 2005). Because forest trees are continuously challenged by changing environmental conditions during their lifetime, adaptive genetic variation in relevant genes and phenotypic plasticity are essential for the long-term adaptation to stressful conditions. Thus, the knowledge of adaptive genetic variation is a basis for future management and conservation strategies of forests (Krutovsky and Neale 2005) and can assist in breeding in combination with traditional phenotypic selection (Neale 2007). Furthermore, the results presented here are a prerequisite for association mapping studies in order to identify genomic regions and even individual nucleotides underlying phenotypic variation. The success of such an approach largely depends on the reasonable selection of candidate genes. This study revealed huge differences in diversity among the investigated candidate genes. Whereas the genes with regulatory function such as the cys-his-zinc finger protein (*CHZFP*) representing a transcription factor show low or moderate SNP variation, genes with a structural function as the ascorbate peroxidase show comparatively high SNP variation.

In the view of the above considerations, we propose to apply the genomic resources developed for beech by the identification and characterization of SNPs in coding and non-coding regions of candidate genes to investigate both the genetic basis of adaptive variation and the population structure of beech at the ultimate level of genetic resolution. In future, there will be the possibility to use whole genome sequencing for these applications. But considering the costs and the possibilities at the moment for non-model organisms, the comparable sequencing of (partial) genes and the identification of SNPs presented in this study is the best available method.

Acknowledgments The study was supported by the Ministry for Science and Culture of Lower Saxony within the network KLIFF—climate impact and adaptation research in Lower Saxony. We thank A. Dolynska, G. Dinkel, and A. Capelle for their technical help and all persons who assisted us doing field work. Furthermore, we thank K. Prinz for valuable comments and discussions.

Conflict of interest The authors declare that they have no conflict of interest.

References

Ammer C, Albrecht L, Borchert H, Brosinger F, Dittmar C, Elling W, Ewald J, Felbermeier B, von Gilsa H, Huss J, Kenk G, Kölling C, Kohnle U, Meyer P, Mosandl R, Moosmayer HU, Palmer S, Reif A, Rehfuess KE, Stimm B (2005) Future suitability of beech (*Fagus sylvatica* L.) in Central Europe: critical remarks concerning a paper of Rennenberg et al. (2004). Allg Forst-u J-Ztg 176:60–67

Baek JM, Han P, Iandolino A, Cook DR (2008) Characterization and comparison of intron structure and alternative splicing between *Medicago truncatula, Populus trichocarpa*, *Arabidopsis* and rice. Plant Mol Biol 67:499–510

Beck EH, Fettig S, Knake C, Hartig K, Bhattarai T (2007) Specific and unspecific responses of plants to cold and drought stress. J Biosci 32:501–510

Boggs JZ, Loewy K, Bibee K, Heschel MS (2010) Phytochromes influence stomatal conductance plasticity in *Arabidopsis thaliana*. Plant Growth Regul 60:77–81

Breathnach R, Benoist C, O'Hare K, Gannon F, Chambon P (1978) Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proc Natl Acad Sci USA 75:4853–4857

Brookes A (1999) The essence of SNPs. Gene 234:177–186

Buiteveld J, Vendramin GG, Leonardi S, Kamer K, Geburek T (2007) Genetic diversity and differentiation in European beech (*Fagus sylvatica* L.) stands varying in management history. For Ecol Manag 247:98–106

Chou PY, Fasman GD (1974) Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. Biochemistry 13:211–222

Derory J, Scotti-Saintagne C, Bertocchi E, Le Dantec L, Graignic N, Jauffres A, Casasoli M, Chancerel E, Bodenes C, Alberto F, Kremer A (2010) Contrasting relations between diversity of candidate genes and variation of bud burst in natural and segregating populations of European oaks. Heredity 105: 401–411

Deschamps S, Campbell MA (2010) Utilization of next-generation sequencing platforms in plant genomics and genetic variant discovery. Mol Breed 25:553–570

Eriksson G (1998) Evolutionary forces influencing variation among populations of *Pinus sylvestris*. Silva Fenn 32:173–184

Eveno E, Collada C, Guevara MA, Leger V, Soto A, Diaz L, Leger P, González-Martínez SC, Cervera MT, Plomion C, Garnier-Gere PH (2008) Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. Mol Biol Evol 25:417–437

Ewing B, Hillier LD, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8:175–185

Frewen BE, Chen THH, Howe GT, Davis J, Rohde A, Boerjan W, Bradshaw HD (2000) Quantitative trait loci and candidate gene mapping of bud set and bud flush in *Populus*. Genetics 154: 837–845

Gailing O, von Wühlisch G (2004) Nuclear markers (AFLPs) and chloroplast microsatellites differ between *Fagus sylvatica* and *F. orientalis*. Silvae Genet 53:105–110

Gailing O, Vornam B, Leinemann L, Finkeldey R (2009) Genetic and genomic approaches to assess adaptive genetic variation in plants: forest trees as a model. Physiol Plant 137:509–519

Gao CX, Han B (2009) Evolutionary and expression study of the aldehyde dehydrogenase (ALDH) gene superfamily in rice (*Oryza sativa*). Gene 431:86–94

Garvin MR, Saitoh K, Gharrett AJ (2010) Application of single nucleotide polymorphisms to non-model species: a technical review. Mol Ecol Resour 10:915–934

Gessler A, Keitel C, Kreuzwieser J, Matyssek R, Seiler W, Rennenberg H (2007) Potential risks for European beech (*Fagus sylvatica* L.) in a changing climate. Trees-Struct Funct 21:1–11

Glenn TC (2011) Field guide to next-generation DNA sequencers. Mol Ecol Resour 11:759–769

González-Martínez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB (2006) DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. Genetics 172:1915–1926

Guo P, Baum M, Grando S, Ceccarelli S, Bai G, Li R, von Korff M, Varshney RK, Graner A, Valkoun J (2009) Differentially expressed genes between drought-tolerant and drought-sensitive barley genotypes in response to drought stress during the reproductive stage. J Exp Bot 60:3531–3544

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acid S 41:95–98

Helyar SJ, Hemmer-Hansen J, Bekkevold D, Taylor MI, Ogden R, Limborg MT, Cariani A, Maes GE, Diopere E, Carvalho GR, Nielsen EE (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. Mol Ecol Resour 11:123–136

Ingvarsson PK (2004) Nucleotide polymorphism and linkage disequilbrium within and among natural populations of European Aspen (*Populus tremula* L., Salicaceae). Genetics 169:945–953

Ingvarsson PK, Garcia MV, Hall D, Luquez V, Jansson S (2006) Clinal variation in phyB2, a candidate gene for day-length-induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*). Genetics 172: 1845–1853

Jimenez JA, Alonso-Ramerez A, Nicolas C (2008) Two cDNA clones (FsDhn1 and FsClo1) up-regulated by ABA are involved in drought responses in *Fagus sylvatica* L. seeds. J Plant Physiol 165:1798–1807

Jones ES, Sullivan H, Bhattramakki D, Smith JSC (2007) A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.). Theor Appl Genet 115:361–371

Kalinowski ST (2002) How many alleles per locus should be used to estimate genetic distances? Heredity 88:62–65

Kim R, Guo JT (2010) Systematic analysis of short internal indels and their impact on protein folding. BMC Struct Biol 10:24

Kraj W, Sztorc A (2009) Genetic structure and variability of phenological forms in the European beech (*Fagus sylvatica* L.). Ann For Sci 66:203

Krutovsky KV, Neale DB (2005) Forest genomics and new molecular genetic approaches to measuring and conserving adaptive genetic diversity in forest trees. In: Geburek T, Turok J (eds) Conservation and management of forest genetic resources in Europe. Arbora Publishers, Zvolen, pp 369–390

Lalagüe H, Fady B, Garnier-Géré P, González-Martínez SC, Lin YC, Oddou-Muratorio S, Sebastiani F, Vendramin GG (2010) Candidate gene variation in common beech (*Fagus sylvatica* L.) along an altitudinal gradient. In: Vinceti B, Neate P (comps.) Conference on "Forest Ecosystem Genomics and Adaptation". San Lorenzo de El Escorial (Madrid), Spain, 9–11 June 2010. Book of abstracts. Bioversity International (Rome, Italy) and INIA (Madrid, Spain), p 242

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 11:2453–2465

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Liu YH, Shi YS, Song YC, Wang TY, Li Y (2010) Characterization of a stress-induced NADP-isocitrate dehydrogenase gene in maize confers salt tolerance in *Arabidopsis*. J Plant Biol 53:107–112

Lu Y, Zhang S, Shah T, Xie C, Hao Z, Li X, Farkhari M, Ribaut JM, Cao M, Rong T, Xu Y (2010) Joint linkage–linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. Proc Natl Acad Sci USA 107:19585–19590

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. Trends Ecol Evol 19:208–216

Muleo R, Morini S, Casano S (2001) Photoregulation of growth and branching of plum shoots: Physiological action of two photosystems. In Vitro Cell Dev-Pl 37:609–617

Müller-Starck G, Starke R (1993) Inheritance of isoenzymes in European beech (*Fagus sylvatica* L.). J Hered 84:291–296

Müller-Starck G, Ziehe M (1991) Genetic variation in populations of *Fagus sylvatica* L., *Quercus robur* L., and *Q. petraea* Liebl. in Germany. In: Müller-Starck G, Ziehe M (eds) Genetic variation in European populations of forest trees. J. D. Sauerländer's Verlag, Frankfurt am Main

Neale DB (2007) Genomics to tree breeding and forest health. Curr Opin Genet Dev 17:539–544

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, Bustamante CD (2008) Genes mirror geography within Europe. Nature 456:98–101

Nyári L (2010) Genetic diversity, differentiation and spatial genetic structure in differently managed adult European beech (*Fagus sylvatica* L.) stands and their regeneration. Forstarchiv 81:156–164

Oddou-Muratorio S, Klein EK, Vendramin GG, Fady B (2011) Spatial vs. temporal effects on demographic and genetic structures: the roles of dispersal, masting and differential mortality on patterns of recruitment in *Fagus sylvatica*. Mol Ecol 20:1997–2010

Olbrich M, Betz G, Gerstner E, Langebartels C, Sandermann H, Ernst D (2005) Transcriptome analysis of ozone-responsive genes in leaves of European beech (*Fagus sylvatica* L.). Plant Biol 7:670–676

Olbrich M, Gerstner E, Bahnweg G, Haberle KH, Matyssek R, Welzl G, Heller W, Ernst D (2010) Transcriptional signatures in leaves of adult European beech trees (*Fagus sylvatica* L.) in an experimentally enhanced free air ozone setting. Environ Pollut 158:977–982

Olson MS, Robertson AL, Takebayashi N, Silim S, Schroeder WR, Tiffin P (2010) Nucleotide diversity and linkage disequilibrium in balsam poplar (*Populus balsamifera*). New Phytol 186:526–536

Pastorelli R, Smulders MJM, Van't Westende WPC, Vosman B, Giannini R, Vettori C, Vendramin GG (2003) Characterization of microsatellite markers in *Fagus sylvatica* L. and *Fagus orientalis* Lipsky. Mol Ecol Notes 3:76–78

Quang ND, Ikeda S, Harada K (2008) Nucleotide variation in *Quercus crispula* Blume. Heredity 101:166–174

Ramanjalu S, Bartels D (2002) Drought- and dessication-induced modulation of gene expression in plants. Plant, Cell Environ 25:141–151

Rennenberg H, Seiler W, Matyssek R, Gessler A, Kreuzwieser J (2004) European beech (*Fagus sylvatica* L.)—a forest tree without future in the south of Central Europe? Allg Forst-u J-Ztg 175:210–224

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics: methods and protocols (Methods in Molecular Biology). Humana Press, Totowa, pp 365–386

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Sathyan P, Newton RJ, Loopstra CA (2005) Genes induced by WDS are differentially expressed in two populations of aleppo pine (*Pinus halepensis*). Tree Genet Genomes 1:166–173

Schlink K (2011) Gene expression profiling in wounded and systemic leaves of *Fagus sylvatica* reveals up-regulation of ethylene and jasmonic acid signalling. Plant Biol 13:445–452

Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, Seeb LW (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organism. Mol Ecol Resour 11:1–8

Street NR, Skogstrom O, Sjodin A, Tucker J, Rodriguez-Acosta M, Nilsson P, Jansson S, Taylor G (2006) The genetics and genomics of the drought response in *Populus*. Plant J 48:321–341

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Vidalis A (2011) Patterns of nucleotide variation and gene-associated SNP analysis in a *Quercus* spp. forest at isocitrate dehydrogenase genes. Ph.D. dissertation. Georg-August-University Göttingen

Vornam B, Decarli N, Gailing O (2004) Spatial distribution of genetic variation in a natural beech stand (*Fagus sylvatica* L.) based on microsatellite markers. Conserv Genet 5:561–570

Vornam B, Gailing O, Finkeldey R, Collada C, Guevera M, Soto Á, de María N, González-Martínez S, Díaz L, Alia R, Aranda I, Climent J, Cervera MT, Goicoechea P, Léger V, Eveno E, Derory J, Garnier-Géré P, Kremer A, Plomion C (2007) Naturally occurring nucleotide diversity in candidate genes for forest tree adaptation: magnitude, distribution and association with quantitative trait variation. GABI—The German Plant Genome Research Program Progress Report 2004–2007, pp 116–120

Vornam B, Gailing O, Derory J, Plomion C, Kremer A, Finkeldey R (2011) Characterisation and natural variation of a *dehydrin* gene in *Quercus petraea* (Matt.) Liebl. Plant Biol 13:881–887

Wachowiak W, Balk PA, Savolainen O (2009) Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold-related candidate genes in Scots pine (*Pinus sylvestris* L.). Tree Genet Genomes 5:117–132

Yoshiura K, Kinoshita A, Ishida T, Ninokata A, Ishikawa T, Kaname T, Bannai M, Tokunaga K, Sonoda S, Komaki R, Ihara M, Saenko VA, Alipov GK, Sekine I, Komatsu K, Takahashi H, Nakashima M, Sosonkina N, Mapendano CK, Ghadami M, Nomura M, Liang DS, Miwa N, Kim DK, Garidkhuu A, Natsume N, Ohta T, Tomita H, Kaneko A, Kikuchi M, Russomando G, Hirayama K, Ishibashi M, Takahashi A, Saitou N, Murray JC, Saito S, Nakamura Y, Niikawa N (2006) A SNP in the ABCC11 gene is the determinant of human earwax type. Nat Genet 38:324–330